

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-325917
 (43)Date of publication of application : 16.12.1997

(51)Int.Cl. G06F 12/16
 G06F 3/06
 G06F 3/06

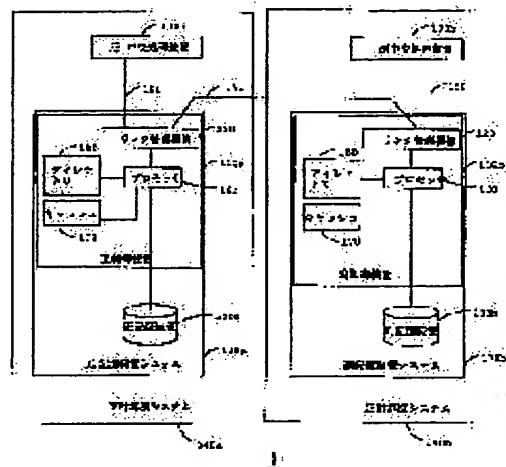
(21)Application number : 08-145562 (71)Applicant : HITACHI LTD
 (22)Date of filing : 07.06.1996 (72)Inventor : YAMAKAMI KENJI
 NAKAMURA KATSUNORI
 YAMAMOTO AKIRA

(54) COMPUTER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a computer system in which whether data are last or not can be recognized by a subordinate storage device when a regular controller can not be used because of any fault or disaster in the case of applying an asynchronous copy system.

SOLUTION: In accordance with a write request from a central processing unit 100, a regular controller 110a transfers only the position information of write object data to a subordinate controller 110b. At the subordinate controller 110b, the position information is held until data are actually written. When a regular storage device 130a can not be accessed because of any fault or disaster, the position information is read out of the subordinate controller 110b so that the presence/absence of erased data, its position or the time can be discriminated.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-325917

(43)公開日 平成9年(1997)12月16日

(51)Int.Cl. [®] G 0 6 F 12/16 3/06	識別記号 3 1 0 3 0 4 3 0 6	序内整理番号 7623 -5B	F I C 0 6 F 12/16 3/06	技術表示箇所 3 1 0 J 3 0 4 F 3 0 6 B
---	---------------------------------	--------------------	------------------------------	---

審査請求 未請求 請求項の数10 O.L (全 14 頁)

(21)出願番号 特願平8-145562	(71)出願人 株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(22)出願日 平成8年(1996)6月7日	(72)発明者 山神 勝司 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
	(72)発明者 中村 勝憲 神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内
	(72)発明者 山本 彰 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
	(74)代理人 弁理士 有近 紳志郎

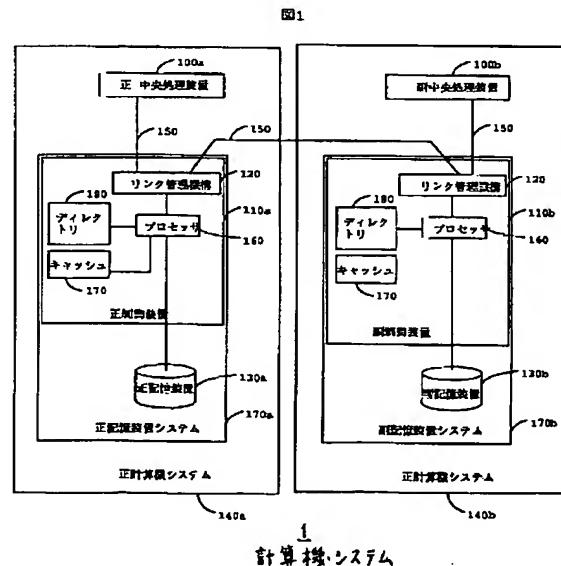
(54)【発明の名称】 計算機システム

(57)【要約】

【課題】 非同期コピー方式を適用した場合において、障害や災害等により正制御装置が使用不能となった時に、データが消失しているかどうかを副記憶装置で認識できる計算機システムを提供する。

【解決手段】 中央処理装置100からのライト要求に対して、正制御装置110aはライト対象データの位置情報のみを副制御装置110bへ転送する。副制御装置110bでは、実際にデータがライトされるまで、前記位置情報を保持しておく。障害や災害により正記憶装置130aがアクセス不能になった場合、副制御装置110bから位置情報を読み出すことにより、消失データの有無およびその位置や時刻を判別する。

【効果】 障害や災害等により正記憶装置が使用不能となった時に、消失データの有無を副制御装置から判断でき、適切に対処することが可能になる。



【特許請求の範囲】

【請求項1】 中央処理装置に接続しあつ記憶装置と制御装置から構成される正記憶装置システムと、その正記憶装置システムに接続する遠隔の副記憶装置システムとから構成される計算機システムにおいて、

前記正記憶装置システム下の制御装置すなわち正制御装置は、前記中央処理装置からライト要求を受けた際、前記副記憶装置へはライトデータは転送せずに、その位置情報を転送した後、前記中央処理装置に対してはライト完了を報告して、後の適当な時期に前記副記憶装置に対するライト処理を実行し、

前記副記憶装置システム下の制御装置すなわち副制御装置は、前記正制御装置から転送された前記位置情報を保持しておき、後に実際のライト処理が実行された時に、前記位置情報を破棄し、

災害等により前記正制御装置が使用不能となったときに、前記副制御装置に格納されている前記位置情報によって、消失したデータの有無を検出可能としたことを特徴とする計算機システム。

【請求項2】 請求項1に記載の計算機システムにおいて、前記正制御装置は、前記位置情報に加えて、前記中央処理装置からライト要求のあった時刻を前記副制御装置に転送することによって、前記中央処理装置のデータ更新順序を、前記副制御装置に記憶させ、

災害等により前記正制御装置が使用不能になったときに、前記副制御装置に格納されている時刻から、回復すべきバックアップファイルを決定し、データ回復を行うことを特徴とする計算機システム。

【請求項3】 請求項1に記載の計算機システムにおいて、前記中央処理装置から前記副制御装置に対して、前記位置情報に対応する領域にアクセスがあった場合に、前記副制御装置は、エラー終了を前記中央処理装置に報告することを特徴とする計算機システム。

【請求項4】 中央処理装置に接続しあつ記憶装置と制御装置から構成される正記憶装置システムと、前記正記憶装置システムに接続する遠隔の副記憶装置システムから構成される計算機システムにおいて、

前記正記憶装置システム下の制御装置すなわち正制御装置と、前記副記憶装置システム下の制御装置すなわち副制御装置は、それぞれが、二重書きの状態として2つの状態、すなわち二重書き一致状態および二重書き不一致状態を管理しており、

前記正制御装置は、前記中央処理装置からライト要求を受けた際、ライト対象の正記憶装置が二重書き一致状態であれば、前記正記憶装置および前記副記憶装置のそれぞれの前記二重書き状態を二重書き不一致状態に遷移させ、前記副記憶装置へはライトデータは転送せずに、前記中央処理装置に対してはライト完了を報告して、後の適当な時期に前記副記憶装置に対するライト処理を実行し、さらに、後に前記副制御装置に対する未更新データ

が全て更新された時に二重書き一致状態に遷移させ、災害等により前記正制御装置が使用不能となった時に、前記副制御装置に格納されている前記二重書き状態によって、データ消失の有無を検出可能としたことを特徴とする計算機システム。

【請求項5】 請求項4に記載の計算機システムにおいて、前記正制御装置および前記副制御装置を、前記二重書き不一致状態に遷移させる際に、前記二重書き不一致状態とした時刻を、前記正制御装置および前記副記憶装置にそれぞれ記録し、

災害等により前記正制御装置が使用不能になった場合に、前記副記憶装置が二重書き不一致状態となつていれば、前記二重書き不一致状態となった時刻に対応するバックアップファイルからデータ回復を行なうことを特徴とする計算機システム。

【請求項6】 請求項2または請求項5に記載の計算機システムにおいて、前記位置付け情報あるいは前記二重書き不一致状態に付随して前記副制御装置に転送される時刻は、ライト処理要求時刻として、前記中央処理装置から与えられることを特徴とする計算機システム。

【請求項7】 請求項2または請求項5に記載の計算機システムにおいて、前記位置付け情報あるいは前記二重書き不一致状態に付随して前記副制御装置に転送される時刻は、前記中央処理装置からライト要求のあった時刻として、前記正制御装置の内部時間を使用することを特徴とする計算機システム。

【請求項8】 請求項2に記載の計算機システムにおいて、前記位置情報は、ライト系アクセスであることが判明したときのみ、前記副制御装置に転送することを特徴とする計算機システム。

【請求項9】 請求項5に記載の計算機システムにおいて、前記二重書き不一致状態は、ライト系アクセスであることが判明したときのみ、前記副制御装置に通知することを特徴とする計算機システム。

【請求項10】 請求項4に記載の記憶装置システムにおいて、前記中央処理装置から、前記二重書き不一致状態である副記憶装置に対してアクセスがあった場合に、前記副制御装置は、エラー終了を前記中央処理装置に報告することを特徴とする計算機システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、中央処理装置に接続しあつ記憶装置と制御装置から構成される正記憶装置システムと、その正記憶装置システムに接続する遠隔の副記憶装置システムとから構成される計算機システムに関する。さらに詳しくは、前記正記憶装置システム下の記憶装置すなわち正記憶装置と、前記副記憶装置システム下の記憶装置すなわち副記憶装置の間で遠隔二重書きを行なう際のデータ保証方法を改良した計算機システムに関する。

【0002】

【従来の技術】中央処理装置を介すことなしに制御装置間でデータ転送を行ない、異なる制御装置間で二重書きを実現する技術、すなわち遠隔二重書きは、U.S.P. 5,155,845に記載されている。この従来技術では、中央処理装置から正／副を指定するコマンドを正制御装置が受けとると、正記憶装置に記憶していたデータを副記憶装置にコピーする。それと同時に、中央処理装置から正制御装置へのライト要求に対して、ライトデータを正記憶装置に格納すると共に、当該ライトデータを副制御装置に転送し、その後、中央処理装置に対してライト完了を報告する。これを同期コピーと呼ぶ。以上の動作により、遠隔二重書きが実現される。遠隔二重書きを行なうことによって、災害や障害等によって、正記憶装置のデータがアクセス不能になった場合に、副記憶装置によって業務を引き継ぐことが可能になる。

【0003】

【発明が解決しようとする課題】上記の同期コピーでは、正制御装置と副制御装置の間の接続距離が数百kmと長い場合には、正制御装置から副制御装置へのデータ転送にかかる時間が長くなるために、中央処理装置に対するアクセス性能が劣化する。この解決のためには、副制御装置に対してライトデータを転送する前にライト完了を中央処理装置に報告しておき、後に副制御装置に対してライトデータを転送する、非同期コピー方式が考えられる。しかし、非同期コピー方式では、ある瞬間に正記憶装置にはデータをライト済みで、副記憶装置にはライト未済である状態が生じる。この時に障害や災害等により正制御装置が使用不能となると、副記憶装置では、データが消失しているかどうかを認識する手段がない。このために、副記憶装置が使用可能かどうかを判断できなくなる問題点がある。そこで、本発明の第一の目的は、非同期コピー方式を適用した場合において、障害や災害等により正制御装置が使用不能となった時に、データが消失しているかどうかを副記憶装置で認識できる計算機システムを提供することにある。

【0004】データが消失していることを認識できた場合、その消失したデータを回復する必要がある。消失したデータを回復する方法としては、ある時点でのデータを定期的にテープにバックアップしておき、必要があればバックアップデータからデータを回復して、その時点までジョブの進行を戻す方法がある。この方法でデータ回復を行なうには、データを消失した時刻が判らなくてはならない。そこで、本発明の第二の目的は、消失したデータが中央処理装置から正記憶装置に対してライトされた時刻を副制御装置で認識できる計算機システムを提供することにある。

【0005】

【課題を解決するための手段】第1の観点では、本発明は、中央処理装置に接続しかつ記憶装置と制御装置から

構成される正記憶装置システムと、その正記憶装置システムに接続する遠隔の副記憶装置システムとから構成される計算機システムにおいて、前記正記憶装置システム下の制御装置すなわち正制御装置は、前記中央処理装置からライト要求を受けた際、前記副記憶装置へはライトデータは転送せずに、その位置情報を転送した後、前記中央処理装置に対してはライト完了を報告して、後の適当な時期に前記副記憶装置に対するライト処理を実行し、前記副記憶装置システム下の制御装置すなわち副制御装置は、前記正制御装置から転送された前記位置情報を保持しておき、後に実際のライト処理が実行された時に、前記位置情報を破棄し、災害等により前記正制御装置が使用不能となったときに、前記副制御装置に格納されている前記位置情報によって、消失したデータの有無を検出可能としたことを特徴とする計算機システムを提供する。上記第1の観点による計算機システムでは、非同期コピー方式であるから、中央処理装置に対するアクセス性能を向上できる。また、正制御装置から転送された位置情報を副制御装置で格納しておくから、障害や災害等により正制御装置が使用不能となった時に、データが消失しているかどうかを、前記位置情報により副記憶装置で認識することが出来る。

【0006】第2の観点では、本発明は、上記構成の計算機システムにおいて、前記正制御装置は、前記位置情報に加えて、前記中央処理装置からライト要求のあった時刻を前記副制御装置に転送することによって、前記中央処理装置のデータ更新順序を、前記副制御装置に記憶させ、災害等により前記正制御装置が使用不能になったときに、前記副制御装置に格納されている時刻から、回復すべきバックアップファイルを決定し、データ回復を行うことを特徴とする計算機システムを提供する。上記第2の観点による計算機システムでは、災害等により前記正制御装置が使用不能になったときに、副制御装置に格納された時刻により、どの時点で正記憶装置と副記憶装置の内容が不一致になったかが判る。よって、副記憶装置システムで業務の引き継ぎを行なう際、適正なバックアップ時刻のデータから回復することが出来る。

【0007】第3の観点では、本発明は、上記構成の計算機システムにおいて、前記中央処理装置から前記副制御装置に対して、前記位置情報に対応する領域にアクセスがあった場合に、前記副制御装置は、エラー終了を前記中央処理装置に報告することを特徴とする計算機システムを提供する。上記第3の観点による計算機システムでは、旧世代のデータにアクセスするのを防止することが出来る。

【0008】第4の観点では、本発明は、中央処理装置に接続しかつ記憶装置と制御装置から構成される正記憶装置システムと、前記正記憶装置システムに接続する遠隔の副記憶装置システムから構成される計算機システムにおいて、前記正記憶装置システム下の制御装置すなわ

ち正制御装置と、前記副記憶装置システム下の制御装置すなわち副制御装置は、それぞれが二重書きの状態として2つの状態すなわち二重書き一致状態および二重書き不一致状態を管理しており、前記正制御装置は、前記中央処理装置からライト要求を受けた際、ライト対象の正記憶装置が二重書き一致状態であれば、前記正記憶装置および前記副記憶装置のそれぞれの前記二重書き状態を二重書き不一致状態に遷移させ、前記副記憶装置へはライトデータは転送せずに、前記中央処理装置に対してはライト完了を報告して、後の適当な時期に前記副記憶装置に対するライト処理を実行し、さらに、後に前記副制御装置に対する未更新データが全て更新された時に二重書き一致状態に遷移させ、災害等により前記正制御装置が使用不能となった時に、前記副制御装置に格納されている前記二重書き状態によって、データ消失の有無を検出可能としたことを特徴とする計算機システムを提供する。上記第4の観点による計算機システムでは、非同期コピー方式であるから、中央処理装置に対するアクセス性能を向上できる。また、二重書き状態として、一致または不一致の状態を各記憶装置毎に、正制御装置および副制御装置で管理するから、障害や災害等により正制御装置が使用不能となった時に、データが消失しているかどうかを、前記二重書き状態により副記憶装置で認識することが出来る。

【0009】第5の観点では、本発明は、上記構成の計算機システムにおいて、前記正制御装置および前記副制御装置を、前記二重書き不一致状態に遷移させる際に、前記二重書き不一致状態とした時刻を、前記正制御装置および前記副記憶装置にそれぞれ記録し、災害等により前記正制御装置が使用不能になった場合に、前記副記憶装置が二重書き不一致状態となつていれば、前記二重書き不一致状態となつた時刻に対応するバックアップファイルからデータ回復を行なうことを特徴とする計算機システムを提供する。上記第5の観点による計算機システムでは、災害等により前記正制御装置が使用不能になつたときに、副制御装置に格納された時刻により、どの時点で正記憶装置と副記憶装置の内容が不一致になつたかが判る。よって、副記憶装置システムで業務の引き継ぎを行なう際、適正なバックアップ時刻のデータから回復することが出来る。

【0010】第6の観点では、本発明は、上記構成の計算機システムにおいて、前記位置付け情報あるいは前記二重書き不一致状態に付随して前記副制御装置に転送される時刻は、ライト処理要求時刻として、前記中央処理装置から与えられることを特徴とする計算機システムを提供する。上記第6の観点による計算機システムでは、外部時間との誤差をなくすことが出来る。

【0011】第7の観点では、本発明は、上記構成の計算機システムにおいて、前記位置付け情報あるいは前記二重書き不一致状態に付随して前記副制御装置に転送さ

れる時刻は、前記中央処理装置からライト要求のあった時刻として、前記正制御装置の内部時間を使用することを特徴とする計算機システムを提供する。上記第7の観点による計算機システムでは、中央処理装置から時刻を与える必要がないため、構成を簡単化できる。

【0012】第8の観点では、本発明は、上記構成の計算機システムにおいて、前記位置情報は、ライト系アクセスであることが判明したときのみ、前記副制御装置に転送することを特徴とする計算機システムを提供する。上記第8の観点による計算機システムでは、ライト系アクセスのときのみ位置情報の転送を行うが、正記憶装置と副記憶装置でデータの不一致が起こるのはライト系アクセスのときのみであるため、無駄な通信を回避できる。

【0013】第9の観点では、本発明は、上記構成の計算機システムにおいて、前記二重書き不一致状態は、ライト系アクセスであることが判明したときのみ、前記副制御装置に通知することを特徴とする計算機システムを提供する。上記第8の観点による計算機システムでは、ライト系アクセスのときのみ二重書き不一致状態の通知を行うが、正記憶装置と副記憶装置でデータの不一致が起こるのはライト系アクセスのときのみであるため、無駄な通信を回避できる。

【0014】第10の観点では、本発明は、上記構成の記憶装置システムにおいて、前記中央処理装置から、前記二重書き不一致状態である副記憶装置に対してアクセスがあった場合に、前記副制御装置は、エラー終了を前記中央処理装置に報告することを特徴とする計算機システムを提供する。上記第10の観点による計算機システムでは、旧世代のデータにアクセスするのを防止することが出来る。

【0015】

【発明の実施の形態】

－第1の実施形態－

図1に、本発明の第1の実施形態にかかる計算機システムの構成を示す。なお、以下の説明では、正／副を符号のa, bにより区別するが、正／副を区別しないときは符号のa, bを省く場合もある。

【0016】この計算機システム1は、正計算機システム140aと、その正計算機システム140aに接続され且つ遠隔にある副計算機システム140bとから構成されている。前記正計算機システム140aは、正中央処理装置100aと、正記憶システム170aとから構成されている。前記副計算機システム140bは、副中央処理装置100bと、正記憶システム170bとから構成されている。前記正記憶システム170aは、正制御装置110aと、正記憶装置130aとから構成されている。前記副記憶システム170bは、副制御装置110bと、副記憶装置130bとから構成されている。

【0017】正／副は、ユーザによって決定されるが、

一般的には、平常時に運用する記憶装置を正とし、それをバックアップする記憶装置を副とする。そして、記憶装置130の正または副によって、便宜上あるいは論理上、制御装置110、中央処理装置100、計算機システム140の正および副を決めることとする。一つの制御装置110配下に、正記憶装置と副記憶装置が混在することもあるが、この場合には、その制御装置110は、正記憶装置に対して正制御装置であり、副記憶装置に対して副制御装置となる。

【0018】制御装置110は、中央処理装置100および他の制御装置110とリンク150で接続されている。リンク管理機構120は、使用すべきリンク150の切替え等を行なう。キャッシュ170は、中央処理装置100からライトされたデータあるいは記憶装置130からリードされたデータを一時的に格納しておくメモリである。中央処理装置100からアクセスのあったデータが、キャッシュ170上に存在していれば記憶装置130へのアクセスは行なわない。プロセッサ160は、記憶装置130、キャッシュ170、中央処理装置100、他の制御装置110間のデータ転送を制御する。ディレクトリ180は、キャッシュ170の管理情報を格納するためのメモリである。プロセッサ160は、ディレクトリ180をアクセスすることによって、ヒットミス判定等を行なう。

【0019】ヒットミス判定とは、中央処理装置100からアクセス要求のあったデータの位置すなわち記憶装置アドレス、データアドレス、データ長から、目的とするデータがキャッシュ170上に存在する（ヒット）か否か（ミス）を判定する処理を言う。

【0020】記憶装置130は、磁気ディスク装置である。但し、磁気テープ装置や光ディスク装置などでもかまわない。図2は、磁気ディスク装置の説明図である。この磁気ディスク装置250は、データを記録する複数個のディスク240と、ディスク240上のデータをリード・ライトするためのヘッド230と、制御装置インターフェース220から構成される。ディスク240が1回転する間にヘッド230がアクセス可能な円状の記録単位をトラック200と呼ぶ。トラック200はディスク240上に複数個存在する。各トラックはセクタ260と呼ばれるある決まった大きさの領域に分割されていて、そのセクタがデータアクセスの最小単位（スロット）となる。あるトラック200がアクセス対象になつた時、そのトラック200をリードライトできる位置にヘッド230を移動する。この動作をシークと言う。このディスク装置250では、複数のヘッド230が同時に移動する。ヘッド230は上から昇順に番号付けされていて、それをヘッド番号と呼ぶ。ディスク240が1回転する間にすべてのヘッド230がアクセス可能なトラック200の集まりをシリンドラ210と呼ぶ。シリンドラ210にはディスク240の内側から昇順に番号付け

されていて、それをシリンドラ番号と呼ぶ。

【0021】ディレクトリ180内には、ヒットミス判定テーブルが格納されている。図3は、ヒットミス判定テーブルの構造例である。このヒットミス判定テーブル340は次の表から構成されている。

装置アドレス対応表300：ディスク装置250毎にエントリを持ち、それぞれのエントリは当該ディスク装置に対応するシリンドラ番号対応表310へのポインタを格納している。

シリンドラ番号対応表310：一つのディスク装置のシリンドラ毎にエントリを持ち、それぞれのエントリは当該シリンドラに対応するトラック番号対応表320へのポインタを格納している。

トラック番号対応表320：一つのシリンドラのトラック毎にエントリを持ち、それぞれのエントリは当該トラックに対応するスロット管理表330へのポインタを格納している。

スロット管理表330：一つのトラックのキャッシュ管理情報であり、次の情報をもつ。

データ有無マップ331：トラック上のどのセクタがキャッシュ170上に存在するかを示す情報である。

キャッシュアドレス332：トラック上のセクタがキャッシュ170上に存在するとき、そのデータが格納されているアドレス情報である。

ポインタ333：キュー管理等を行なう時に使用する情報である。

時刻334：当該スロットに対して正記憶装置130aにおいてライトのあった時刻を副制御装置側110b側で管理するための情報である。

RVOLダーティ状態335：当該スロットが副記憶装置130bに対して未更新かどうかを示す情報である。

トラックアドレス336：当該トラックのアドレスである。

ヘッドアドレス337：当該トラックにアクセスするヘッドのアドレスである。

【0022】中央処理装置100から指定された位置付け情報すなわちアクセス対象の記憶装置アドレス、シリンドラ番号、トラック番号を基に、装置アドレス対応表300、シリンドラ番号対応表310、トラック番号対応表320、スロット管理表330をたどり、データ有無マップ331の対応するビットがオンであればヒットと判定し、オフであればミスと判定する。さらに、データ転送を行なう場合には、キャッシュアドレス332に目的のセクタに対応するオフセットを加えることによって、転送対象となるキャッシュアドレスを算出する。

【0023】非同期コピーによる遠隔二重書きを実現するために、副記憶装置130bへの未更新のデータ（これをRVOLダーティデータという）を管理するRVOLダーティ管理表をディレクトリ180にもつ。

【0024】図4は、RVOLダーティ管理表の構成図であ

る。このRVOLダーティ管理表400は、記憶装置VOL1, VOL2, …毎のエントリを有している。一つのエントリには、RVOL情報401と、RVOLダーティリストヘッド402とが格納されている。RVOL情報401には、当該記憶装置を正としたときの副制御装置110bアドレスおよび副記憶装置130bアドレスを保持する。当該記憶装置130が正／副のペアを組んでいない場合は特殊な値が格納される。RVOLダーティリストヘッド402は、未更新のスロットを管理するためのRVOLダーティリスト410を指す。未更新のデータが存在しない場合にはNULL値が格納される。RVOLダーティリスト410は、スロット管理表330内のポインタ333を用いて未更新スロットのスロット管理表330をリスト構造にしたものである。

【0025】RVOLダーティ管理表400の登録は、中央処理装置100から正記憶装置130aへのライト系アクセス時もしくはライト系アクセスと判明した時に実行される。中央処理装置100は、定められたプロトコルに従って、正記憶装置130aへアクセスする。そのプロトコルについては様々なもののが存在するが、例えばメインフレームを中心とした大型計算機システムでは、CKDプロトコルが一般的に使用されている。このCKDプロトコルでは、データアクセスコマンドに先だって、位置付け情報の他に、アクセス形態（リードまたはライト系アクセス）を表す情報を含む位置付けコマンドが発行される。従って、ライト系アクセスかどうかは、位置付けコマンドを受領した時点で判明するため、この時にRVOLダーティ管理表400の登録を行なう。また、ワークステーションやパソコンでは、データアクセスコマンドに位置付け情報を含むFBAプロトコルが用いられている。このFBAプロトコルの場合には、データアクセスコマンド受領後、RVOLダーティ管理表400の登録を行なう。

【0026】図5は、RVOLダーティ管理表400の登録処理のフローチャートである。ステップ510では、上述のようにライト系アクセスか否かを判定し、ライト系アクセスでなければ処理を終了し、ライト系アクセスであることが判明すればステップ520へ進む。ステップ520では、アクセス対象の記憶装置130のアドレスからRVOLダーティ管理表400のエントリを求め、当該エントリのRVOL情報401を参照することによって、アクセス対象の記憶装置130が正／副のペアを形成しているかどうかをチェックする。ペアを形成していないければ処理を終了し、ペアを形成していればステップ530へ進む。ステップ530では、RVOL情報401から副制御装置110bのアドレスを求めて、その副制御装置110bに対して位置付け情報を転送する。これにより、副記憶装置130bに未反映のデータが存在することを副制御装置110bに通知することが出来る。ステップ540では、RVOLダーティ管理表400にライト対象ス

ロットを登録する。すなわち、RVOLダーティリストヘッダ402の指すスロット管理表330の前に、新規のスロット管理表330を挿入する。

【0027】なお、前記ステップ530において、副制御装置110bに位置付け情報を転送する処理は、見かけ上のオーバヘッドを削減するために、中央処理装置からライトデータを受けとる処理と並行して行なうのが良い。例えばCKDプロトコルで説明すると、位置付けコマンドにより位置付け情報を受領した後、子ジョブを生成して位置付け情報転送処理を実行させておき、自ジョブでは位置付けコマンドにチェインするライトコマンドの処理を実行する。自ジョブでは、当該コマンドチェインが完了した後、子ジョブの完了を待ってから中央処理装置100に対してライト処理完了を報告する。

【0028】また、前記ステップ530において、副制御装置110bに対して、位置付け情報の他に、現在の時刻を転送してもよい。これにより、どの時点でデータが消失したかを副制御装置110b側で後に調べることが出来るようになる。さらに、現在の時刻を転送する際、正制御装置110aの内部時間を転送してもよいが、外部時間との誤差をなくすために、例えば中央処理装置100からの位置付けコマンドパラメータ内に時刻を格納しておき、その時刻を転送してもよい。例えば図6に示すRVOLダーティ情報600により、位置付け情報および時刻は、正制御装置110aから副制御装置110bへ転送される。このRVOLダーティ情報600は、RVOLダーティ情報の転送であることを表すコマンドコード601と、副記憶装置アドレス610と、ダーティデータ位置情報620と、時刻630とから構成される。

【0029】図7は、RVOLダーティ管理表400の削除処理のフローチャートである。なお、効率的な非同期コピー方式の遠隔二重書きを実現するために、実装している正記憶装置130a対応にRVOLライトを専属に実行する非同期二重書きジョブが存在するものとする。この非同期二重書きジョブは、一定周期毎に、ランとスリープとを繰り返すジョブで、ラン中にRVOLダーティ管理表400の削除処理を実行する。ステップ710では、実行中の非同期二重書きジョブに対応する正記憶装置130aのRVOLダーティ管理表400のRVOLダーティリストヘッダ402を参照し、その値がNULLならばRVOLダーティデータは存在しないので、処理を終了する。RVOLダーティリストヘッダ402の値がNULLでないならば、RVOLダーティデータが存在するので、ステップ720へ進む。ステップ720では、RVOLダーティ管理表400のRVOL情報401と当該スロット管理表330の情報とから副記憶装置130bに対するライトコマンドを生成し、RVOLライト処理を実行する。なお、ライトコマンドの生成方法については例えばUSP555845に記載されている。ステップ730では、ライト完了したスロット管理表330をRVOLダーティリスト410から

削除する。処理が完了すると、しばらくスリープする。

【0030】以上が正制御装置110a側の処理である。次に、副制御装置110b側の処理を説明する。

【0031】副制御装置110bの主な処理は、RVOLデータ情報600が正制御装置110aから転送されてきてから、当該データがライトされるまでの間、RVOLデータ情報600を保持しておくことである。この処理を実現するため、副制御装置110bでも、図4に示したRVOLデータ管理表400と基本的に同一のテーブル構造をもつ。RVOLデータリスト410に接続されたスロット管理表330が、副記憶装置130bへ未反映のスロットを表す。なお、副制御装置110bでは、RVOL情報401は、使用されない。

【0032】RVOLデータ管理表400への登録は、正制御装置110aからRVOLデータ情報600が転送されてきた時に実行される。すなわち、RVOLデータ情報600の転送かどうかをコマンドコード601によって判定し、RVOLデータ情報600の転送であれば、その記憶装置アドレス610から、対応する副記憶装置130bのRVOLデータ管理表400およびRVOLデータリスト410を求める。また、データ位置情報620からヒットミス判定を実行し、対象となるスロット管理表330を得る。そして、そのスロット管理表330を前記RVOLデータリスト410に登録すると共に、そのスロット管理表330のRVOLデータ状態335を“データあり”とする。

【0033】スロット管理表330の削除は、正制御装置110aからの対象スロットに対するライトを実行した時に実行される。すなわち、正制御装置110aからのライト要求に対して、位置付け情報を基にヒットミス判定を行ない、得られたスロット管理表330のRVOLデータ状態335を参照し、“データあり”ならば、ライト処理を実行した後、RVOLデータリスト410から当該スロット管理表330を削除し、RVOLデータ情報335を“データなし”に変更する。

【0034】以上によって、副記憶装置130bに対する未更新データの位置情報およびそのデータが未更新となった時刻が、正制御装置110aおよび副制御装置110bでそれぞれ保持されることになる。

【0035】図8に、消失データの位置情報、消失時刻を記録するための消失データリストを示す。この消失データリスト1100は、記憶装置アドレス1110と、消失データ数1120と、消失データ毎のデータアドレス1130および消失した時刻1140とから構成される。

【0036】なお、前記消失データ数が所定の閾値を越えると、記憶装置130全体が無効になり、記憶装置130全体を回復するようにしてもよい。こうすると、消失データリスト1100が膨大になるのを防ぐことが出来る。前記閾値は、保守員などから設定できるのが好ま

しい。

【0037】副制御装置110bは、中央処理装置100からの消失データリスト1100の読み出し要求に対して、消失データリスト1100を次の手順により作成する。

(1) 要求のあった記憶装置アドレスから、対応するRVOLデータ管理表400を検索する。

(2) 検索したRVOLデータ管理表400のRVOLデータリストヘッド402の値がNULLなら、消失データは存在しないので、消失データ数1120に“0”を記録する。

(3) RVOLデータリストヘッド402の値がRVOLデータリスト410を指しているなら、そのRVOLデータリスト410に接続しているスロット管理表330のトラックアドレス336、ヘッドアドレス337をデータアドレス1130に格納し、時刻334を時刻1140に格納する。これを、RVOLデータリスト410をたどりながら、繰り返し、消失データ数をカウントしていく。

(4) カウントしていた消失データ数を消失データ数1120に格納する。

(5) 指定された記憶装置アドレスを記憶装置アドレス1110に格納する。

【0038】以上により消失データリスト1100が完成すると、副制御装置110bは、中央処理装置100に消失データリスト1100を返送する。そこで、保守員は、消失データリスト1100を読み出すことにより、消失データの有無を調べることが出来る。そして、データが消失していた場合には、データアドレス1130からデータを回復することが出来る。また、時刻1140の最も古いものを検索して、その世代のバックアップファイルからデータを回復することも可能となる。

【0039】図9は、正制御装置がアクセス不可能になっている場合の副制御装置110bにおけるアクセス可否判定処理800を表すフローチャートである。ステップ810では、アクセス対象スロットのヒットミス判定を実行して、スロット管理表330を得る。ステップ820では、前記スロット管理表330のRVOLデータ状態335から、当該スロットのデータが副記憶装置130bに未反映かどうかを判定する。未反映であれば、データが消失しているため、ステップ830へ進む。未反映でなければ、データが消失していないため、ステップ840へ進む。

【0040】ステップ830では、中央処理装置100に対して異常終了を報告する。ステップ840では、アクセスを許可する。

【0041】以上で第1の実施形態の説明を終るが、ここで第1の実施形態の要点をまとめると次のようになる。

(1) 正制御装置110aからライト処理を実際に行う

前に、副制御装置110bへ位置付け情報および時刻を転送しておく。副制御装置110bは、実際にライトが行なわれるまで、前記位置付け情報と前記時刻を保持しておく。

(2) 障害、災害等により正記憶装置130aがアクセス不能となった場合は、副制御装置110bから消失データの有無を判断する。消失データが有った場合には、その消失データの時刻から、回復すべきバックアップファイルを決定する。

(3) 消失データに対してアクセスがあった場合は、アクセスを許可せず、異常終了を報告する。

【0042】従来の非同期コピー方式の遠隔二重書きでは、消失したデータの有無を判断できないため、消失データが存在するにもかかわらずデータの回復を行なわなかっただ場合には、データ化けが発生する。一方、消失データが存在しないにもかかわらずデータの回復を行なった場合には、回復作業が無駄になり、データが数世代前に戻るため、極めてコストがかかる。また、記憶装置130全体を回復しなくてはならず、効率が悪い。これに對して、上記第1の実施形態による非同期コピー方式の遠隔二重書きでは、副側のサイトで運用を開始する前に、まず正記憶装置130aと副記憶装置130bの内容が一致しているかどうかをチェックして、不一致であれば適切な時刻のバックアップファイルからデータ回復してから運用を開始する。従って、前記データ化けや無駄なデータ回復を回避することが出来る。以上により、災害や障害等によって正記憶装置110aに対するアクセスが不可能になった場合でも、副記憶装置130bを使用して業務を副側のサイトで引き継ぐことができ、しかも、システム停止時間を最小にできる。

【0043】－第2の実施形態－

前記第1の実施形態では、中央処理装置100からの1回のライト要求毎(CCチェイン毎)に副制御装置110bに位置付け情報を転送していたため、正制御装置110aと副制御装置110bの間の通信回数が多くなり、性能が低下する。これを解決するため、第2の実施形態では、正記憶装置130aと副記憶装置130bの内容が一致あるいは不一致のときのみ、そのことを副制御装置110bに通知し、正制御装置110aと副制御装置110bの間の通信回数を削減する。

【0044】記憶装置130は、二重書き一致または二重書き不一致の2状態をとる。ここで、二重書き一致とは、正記憶装置130aと副記憶装置130bの内容が一致した状態であることを示す。また、二重書き不一致とは、正記憶装置130aと副記憶装置130bの内容が不一致の状態であることを示す。これらの二重書き状態は、正制御装置130aおよび副制御装置130bのRVOLダーティ管理表400のRVOL情報401に格納しておく。また、二重書き不一致となった時刻も、RVOL情報401に格納しておく。

【0045】図10は、二重書き一致から二重書き不一致に状態遷移させる不一致化状態変更処理900のフローチャートである。ステップ910では、ライト系アクセスかどうかをチェックする。ライト系アクセス以外であれば、二重書き状態の変更は発生しないので、処理を終了する。ライト系アクセスであれば、ステップ920へ進む。ステップ920では、アクセス対象の記憶装置130が正／副のペアを形成しているかどうかをRVOL情報401から判定する。ペアを形成していないければ、二重書き状態の変更は発生しないので、処理を終了する。ペアを形成していた場合は、ステップ930へ進む。ステップ930では、現在の二重書き状態が二重書き一致かどうかをRVOL情報401から判断する。二重書き不一致であれば、二重書き状態の変更は発生しないので、処理を終了する。二重書き一致であれば、ステップ940へ進む。ステップ940では、副制御装置110bに対して、二重書き不一致となったことを通知する。効率の観点から、この通知の処理と、中央処理装置100からのライト処理とを、並行して行なうのが好ましい。ステップ950では、RVOL情報401を二重書き不一致に状態変更する。

【0046】なお、上記不一致化状態変更処理900で、二重書き不一致の通知だけでなく、その時刻を副制御装置110bに通知してもよい。この場合、中央処理装置100からライト系アクセスを受けた時刻を副制御装置110bへ転送し、かつRVOL情報401にも当該時刻を記録する。なお、この時刻は、正制御装置110aの内部時刻でも良いし、中央処理装置100から与えられる時刻でも良い。

【0047】図11は、二重書き不一致から二重書き一致に状態遷移させる一致化状態変更処理1000のフローチャートである。この一致化状態変更処理1000は、第1の実施形態で説明した非同期二重書きジョブが副記憶装置130bに未更新データをライトした時に実行される。ステップ1000では、第1の実施形態で説明したように未更新データを副記憶装置130bに対してライトする。ステップ1010では、第1の実施形態で説明したようにライト完了したスロット管理表330をRVOLダーティリスト410から外す。ステップ1020では、RVOLダーティリスト410をたどることにより、副記憶装置130bに対する未更新データが存在するかどうかを判断する。RVOLダーティリスト410にスロット管理表330が未だ接続されていれば、二重書き不一致のままであるので、処理を終了する。RVOLダーティリスト410にスロット管理表330が接続されていなければ、二重書き不一致が解消されたので、ステップ1030へ進む。ステップ1030では、二重書き一致となったことを副制御装置110bに通知する。ステップ1040では、RVOL情報401を二重書き一致に状態変更する。

【0048】図12は、正制御装置110aから副制御装置110bに対する状態変更の通知で使用される状態変更通知情報の例示図である。この状態変更通知情報1300は、状態変更通知コマンドであることを示すコマンドコード1301と、対象となる副記憶装置アドレス1310と、状態変更を行なった時刻1320と、新状態のコードを表す状態コード1330とが格納される。

【0049】以上が正制御装置110a側の処理である。次に、副制御装置110b側の処理を説明する。

【0050】副制御装置110bでは、状態変更通知情報1300を受領すると、対象となる副記憶装置130bに対応したRVOL情報401を状態コード1330で指定された状態に変更する。なお、副制御装置110bでは、RVOLダーティリスト410は使用しない。

【0051】図13に、記憶装置130の二重書き状態を中央処理装置100に通知するための正副状態情報1200を示す。この正副状態情報1200は、記憶装置アドレス1210と、二重書き一致あるいは二重書き不一致を表す二重書き状態1220と、二重書き不一致となった時刻を表す状態変更時刻1230とから構成される。

【0052】副制御装置110bは、中央処理装置100からの正副状態情報の読み出し要求に対して、正副状態情報1200を次の手順により作成する。

(1) 指定された記憶装置アドレスから、対応するRVOLダーティ管理表400を検索し、RVOL情報401から、二重書き状態を得て、二重書き状態1220へ格納する。

(2) 二重書き不一致であれば、RVOL情報401に記録された時刻を状態変更時刻1230へ格納する。

以上により正副状態情報1200が完成すると、副制御装置110bは、中央処理装置100に正副状態情報1200を返送する。そこで、保守員は、正副状態情報1200を読み出すことにより、二重書き状態を確認することが出来る。そして、二重書き不一致であれば、不一致となった時刻から最も世代の新しいバックアップファイルを選び出し、データ回復を行なうことが出来る。この第2の実施形態の方法は、個々のデータ回復ができないあるいは不要な場合に有用である。

【0053】中央処理装置100から二重書き不一致の副記憶装置130bに対してアクセス要求があったら、その副制御装置110bは、要求のあった記憶装置アドレスから対応するRVOLダーティ管理表400を求め、そのRVOL情報401から、当該記憶装置130が正／副のペアを組んでいるかどうか、および、正／副のペアを組んでいる場合には二重書き不一致かどうかを調べる。そして、二重書き不一致の場合は、旧世代のデータを転送するのを防ぐために、アクセスを拒否する。

【0054】以上で第2の実施形態の説明を終るが、ここで第2の実施形態の要点をまとめると次のようにな

る。

(1) 正および副記憶装置130の内容の一致／不一致を表す二重書き状態を制御装置110に保持する。

(2) 二重書き状態は、副記憶装置130bに未更新データができる時に二重書き不一致となり、未更新データがなくなった時に二重書き一致となる。

(3) 障害や災害により正記憶装置130aへのアクセスが不可能となった場合は、副制御装置110bから二重書き状態を読み出し、二重書き不一致となった時刻から最も近い世代のバックアップファイルからデータを回復することができる。

(4) 二重書き不一致となっている副記憶装置130bに対する中央処理装置100からのアクセスは拒否される。

【0055】上記第2の実施形態では、副側のサイトで運用を開始する前に、保守員は、正記憶装置130aと副記憶装置130bの内容が一致しているか不一致なのかをチェックする。そして、不一致であれば、適切な時期のバックアップファイルからデータを回復してから、運用を開始する。以上により、災害や障害等によって正記憶装置110aに対するアクセスが不可能になった場合でも、副記憶装置130bを使用して業務を副側のサイトで引き継ぐことができ、しかも、システム停止時間を最小にできる。

【0056】-第1の実施形態と第2の実施形態の差異-

(1) 第1の実施形態では、中央処理装置100からのライト処理の応答時間が劣化する可能性がある。一方、第2の実施形態では、応答性能劣化はほとんどない。

(2) 障害、災害後の消失データの回復では、第1の実施形態では消失したデータ単位に回復可能なのにに対して、第2の実施形態では、記憶装置全体を回復する必要がある。

【0057】

【発明の効果】本発明の計算機システムによれば、副記憶装置に対する未更新データができると、その位置付け情報あるいは正副の二重書き状態が不一致であることを、正制御装置から副制御装置に通知するので、障害や災害等により正記憶装置が使用不能となった時に、消失データの有無を副制御装置から判断でき、適切に対処することが可能になる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態にかかる計算機システムの構成図である。

【図2】磁気ディスク装置の構成図である。

【図3】ヒットミス判定テーブルの構成図である。

【図4】RVOLダーティ管理表の構成図である。

【図5】RVOLダーティ管理表の登録処理を表すフローチャートである。

【図6】RVOLダーティ情報の構成図である。

(10)

特開平9-325917

【図7】RVOLデータ管理表の削除処理を表すフローチャートである。

【図8】消失データリストの構成図である。

【図9】アクセス可否判定処理を表すフローチャートである。

【図10】不一致化状態変更処理を表すフローチャートである。

【図11】一致化状態変更処理を表すフローチャートである。

【図12】状態変更通知情報の構成図である。

【図13】正副状態情報の構成図である。

【符号の説明】

100…中央処理装置

110…制御装置

130…記憶装置

160…プロセッサ

170…キャッシュ

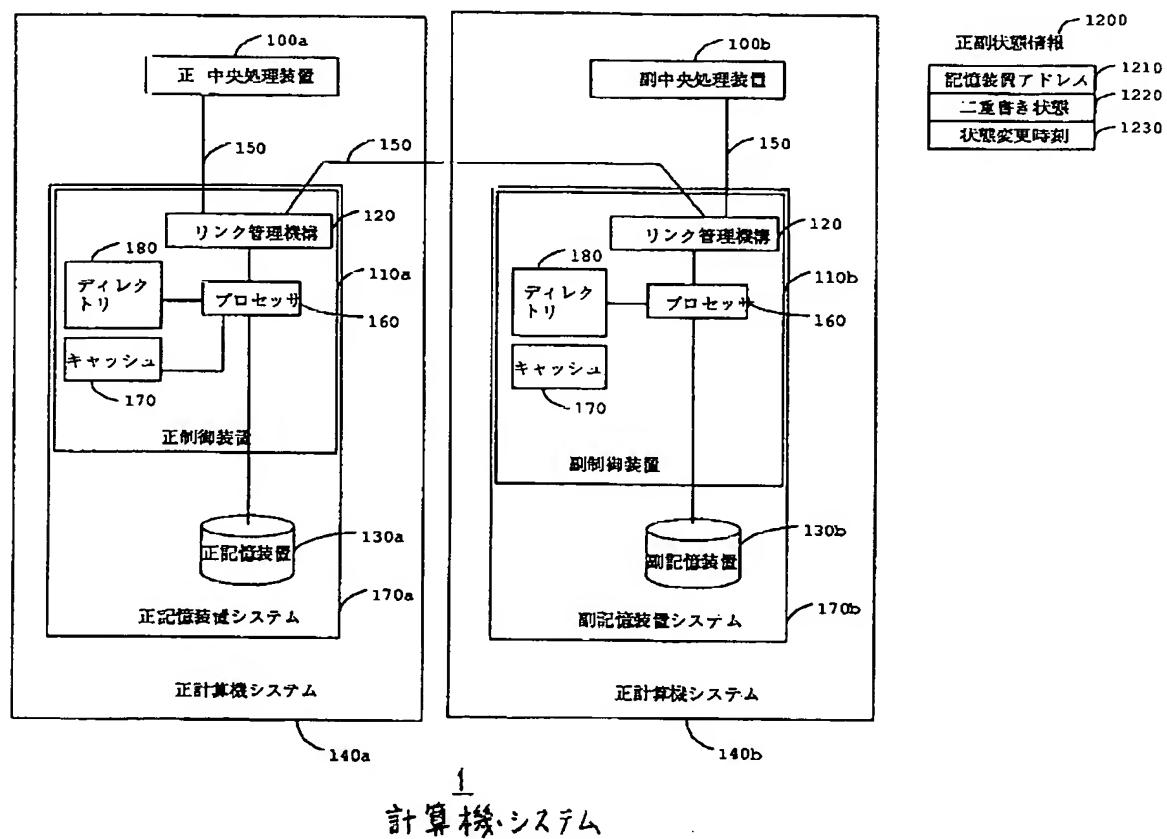
180…ディレクトリ

【図1】

図1

【図13】

図13 正副状態情報



【図6】

図6

【図12】

図12

状態変更通知情報
1300

コマンドコード	副記憶装置アドレス	データ位置情報	時刻
60	610	620	630

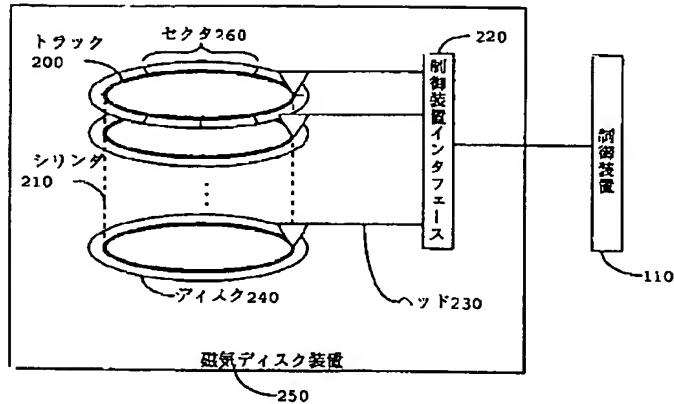
RVOLデータ情報 600

(11)

特開平9-325917

【図2】

図2



【図8】

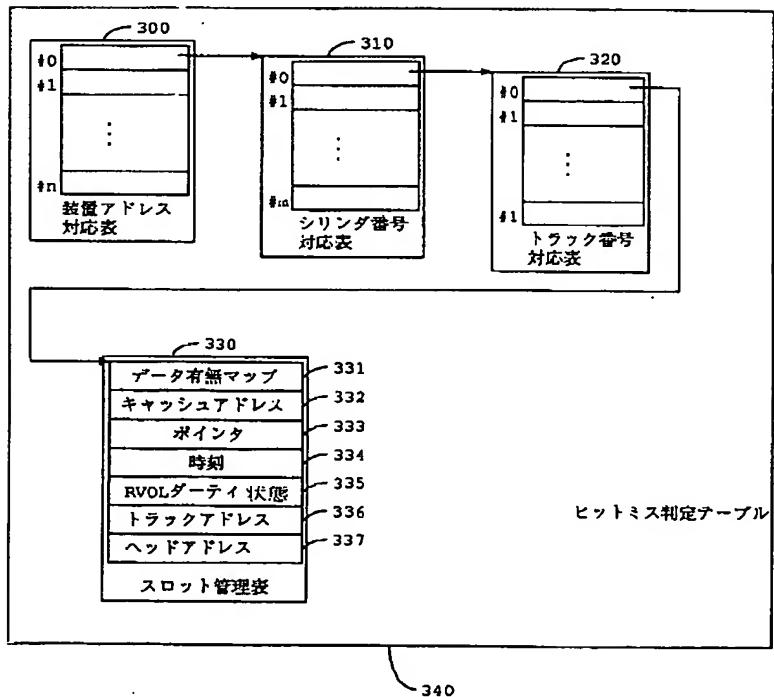
図8

消失データリスト
1100

記録装置アドレス	1110
消失データ数	1120
データアドレス	1130
時刻	1140
:	:
データアドレス	1130
時刻	1140

【図3】

図3

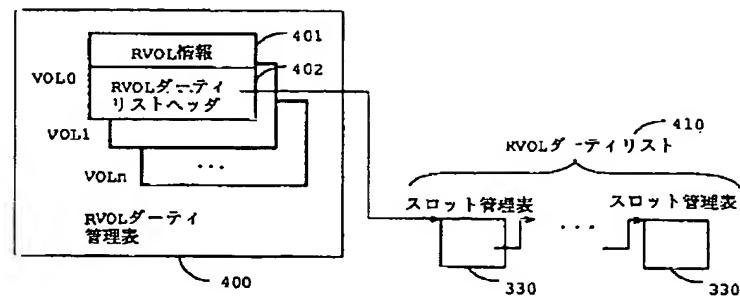


(12)

特開平9-325917

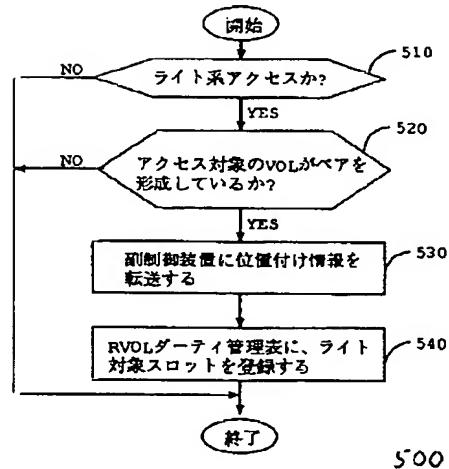
【図4】

図4



【図5】

図5



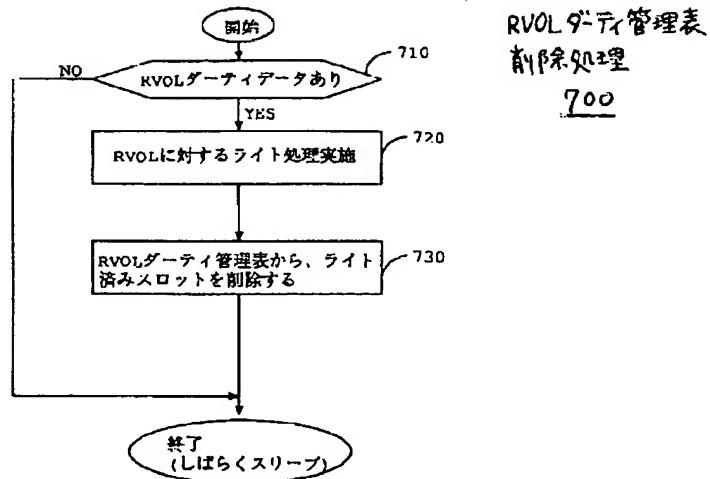
RVOLデータ管理表登録処理

(13)

特開平9-325917

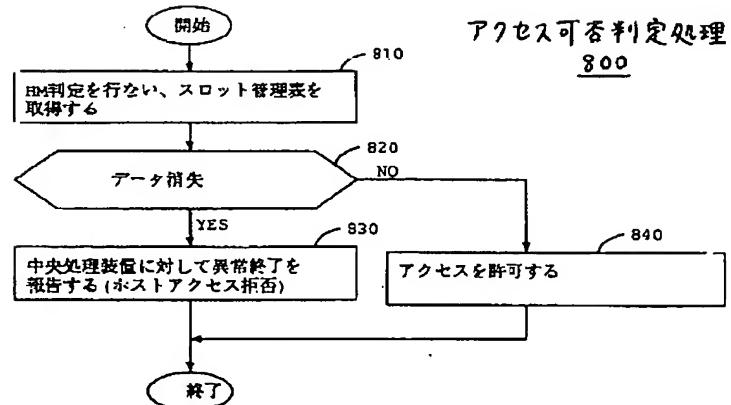
【図7】

図7



【図9】

図9

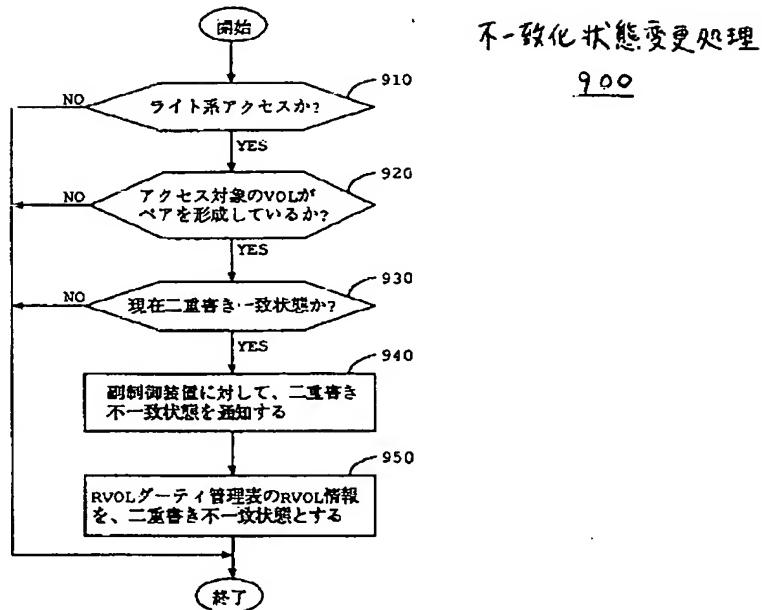


(14)

特開平9-325917

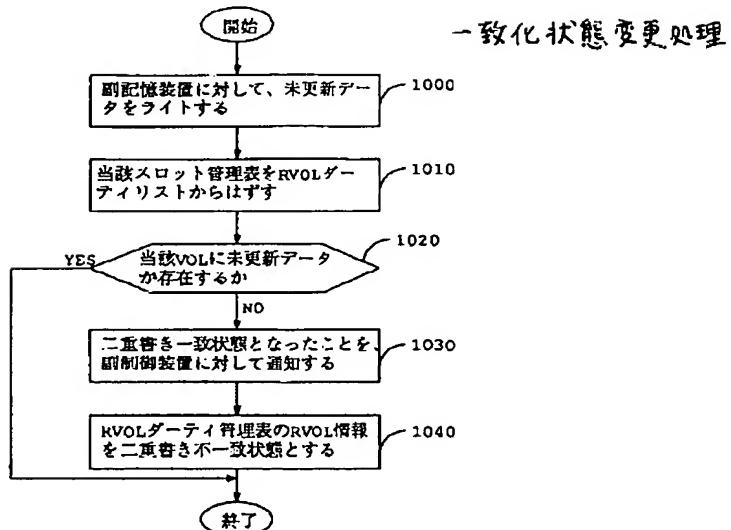
【図10】

図10



【図11】

図11



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.